Data Protection for Research:

When do data protection laws apply to research activities?

When will data protection laws apply to my research? Do I need to think about data protection for my project?

Data protection laws apply to **any processing of personal data** carried out by the University and this will include processing in the course of research activities.

What does "processing" cover?

Processing is any action that is taken with someone's personal data. This starts with collecting personal data or creating a record of information about someone and continues until you no longer need the information and it's been securely destroyed. Just holding the data counts as processing, even if you aren't actively doing anything else with it, so personal data that you have archived or is just retained in our files is still being processed and is still covered by data protection laws. Collecting, sharing or disclosing personal data, combining or matching it with other data, analysing and using it, organising and restructuring it are all kinds of processing activities.

What is personal data?

The **UK GDPR definition** of personal data is:

"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"

An individual is **'identified' or 'identifiable'** if you can distinguish them from other individuals. A name is perhaps the most common means of identifying someone and most of the data held by the University is linked to a name or a unique identifier (staff, student, applicant or research participant number). Corporate email addresses can also directly identify an individual as they are unique identifiers. However, you don't have to know someone's name to distinguish them from other members of a group and identify them. For example: *The elderly man who lives at 15 Purple Street and drives a Porsche Cayenne.*

A combination of identifiers may be needed to identify an individual.

If you are collecting a dataset or conducting a survey which includes some identifiers like name or email address, then it isn't just the name and email address that are personal data. All the data from or about the individuals which is linked to those identifiers will be considered personal data, including any opinions they have expressed or behaviours or experiences they have described.

The UK GDPR specifically refers to **online identifiers** which include IP addresses, cookie identifiers, RFID tags, MAC addresses, account handles and device fingerprints. The ICO advises:

> *"When assessing if an individual is identifiable, you must consider whether online identifiers, on their own or in combination with other information that may be available to those processing the data, may be used to distinguish one user from another, possibly by the creation of profiles of the individuals to identify them. This may be either as a named individual or simply as a unique user of electronic communications and other internet services who may be distinguished from other users."*

If, for instance, you are collecting IP addresses as part of a survey, or the survey tool that you are using automatically collects these and you will have access to them, then the survey responses will be personal data.

It’s important to remember that the information that is collected or created for a research project or is otherwise held by the University may **indirectly identify** an individual and therefore can still be personal data. If so, this means that the information is subject to data protection law The ICO advises:

> *“If you cannot identify an individual directly from the information that you are processing (for example where all identifiers have been removed) an individual may still be identifiable by other means. This may be from information you already hold, or information that you need to obtain from another source. Similarly, a third party could use information you process and combine it with other information available to them. You must carefully consider all of the means that any party is reasonably likely to use to identify that individual…. The key point of indirect identifiability is when information is combined with other information that then distinguishes and allows for the identification of an individual.”*

Examples of information that could allow an individual to be indirectly identified:

- car registration number and/or VIN;
- national insurance number;
- passport number;
- telephone number;
- body measurements and health data; or
- a combination of significant criteria (e.g. age, occupation, place of residence).

Researchers often collect and use the **postcodes** of participants in studies. The final two digits are the part of the postcode that represent a street, part of a street, a single address, or a group of properties, so the use of full postcodes could potentially identify an individual household or a small number of households. The first half of the postcode and the first digit of the second part may suffice for research purposes without the risk of identifying a single address.

Personal data may be collected and stored in a range of different **media** and can take the form of video or audio recordings, photographs and other images such as x-rays.

In addition to your “research data”, you may also be processing the personal data of staff at collaborative partner organisations, research students, PPI contacts etc. This may be limited to names and contact details, but may also include expenses claims, personal opinions, CVs etc. Data protection laws apply to the processing of this personal data too.

Information about a **deceased person** is not considered to be personal data and is not, therefore, subject to UK data protection law. Care and consideration should be given to the data of living relatives and other individuals in relation data about deceased persons as they may be data subjects.

Information about companies or public authorities is not personal data. However, information about individuals acting as sole traders, employees, partners and company directors where they are individually identifiable and the information relates to them as an individual may constitute personal data.

What about pseudonymised and anonymised data?

If the data is fully **anonymised**, then data protection law doesn’t apply because we aren’t processing data that relates to an identified or identifiable individual. This can be beneficial to data subjects and means that data protection compliance measures are not required, but in practice, anonymisation can be difficult to achieve.

The ICO guidance includes a warning:

> *“… you should exercise caution when attempting to anonymise personal data. Organisations frequently refer to personal data sets as having been ‘anonymised’ when, in fact, this is not the case. You should therefore ensure that any treatments or approaches you take truly anonymise personal data. There is a clear risk that you may disregard the terms of the UK GDPR in the mistaken belief that you are not processing personal data.*
>
> *In order to be truly anonymised under the UK GDPR, you must strip personal data of sufficient elements that mean the individual can no longer be identified. However, if you could at any point use any reasonably available means to re-identify the individuals to which the data refers, that data will not have been*

> *effectively anonymised but will have merely been pseudonymised. This means that despite your attempt at anonymisation you will continue to be processing personal data."*

Stripping out the direct identifiers from a data set may not be enough to fully anonymise the data. You will need to think about whether there is still enough information to identify an individual, perhaps from combining or matching the data with other information or knowledge held by the University, collaborative partners, or publicly available.

We can't say that data protection doesn't apply unless we are sure that the data really is anonymised.

When you anonymise personal data, you are still processing personal data up to and at that point. Only after the anonymisation process is complete have you finished processing personal data.

Anonymised data may not fulfil your research purposes. For example, if you need to track individuals in a longitudinal study, then aggregated or anonymous data would clearly make the research impossible.

Pseudonymisation refers to a technique that replaces or removes information in a data set that identifies an individual (e.g. replacing names and other identifiers with a reference number). Pseudonymisation means that people are not identifiable from the dataset itself. However, they are still identifiable by referring to other, separately held information. It doesn't matter that you don't intend to do this. It doesn't matter that you haven't got access to the key. The data is still personal data and data protection law applies.

Pseudonymisation does, however, lessen the security risks associated with the data because you need access to the key to make the identification, so where you can use pseudonymisation, this is a good security measure to deploy.

Pseudonymisation is a helpful technique where you need to collect further information about the individuals e.g. another phase of the research, longitudinal study, further sampling. Researchers may receive a pseudonymised data set from another organisation, with the possibility of requesting further data relating to those data subjects at a later time.

If the process of pseudonymisation is taking place at the University (e.g. to enable more secure data sharing with collaborative partners), you must ensure that they key is stored in a separate location to the pseudonymised data.

Using University student number, NHS number, or other sets of unique identifiers without names is not pseudonymisation. It would still be possible to identify the individual via a number of routes. A single key used only for the research project would need to replace these commonly used unique identifiers.

Where it is feasible to anonymise or pseudonymise personal data, you should ensure that you anonymise or pseudonymise data at the earliest possible opportunity, ideally prior to using the data for research purposes.

What is special category personal data?

The UK GDPR defines special category data as:

- personal data revealing **racial or ethnic origin**;
- personal data revealing **political opinions**;
- personal data revealing **religious or philosophical beliefs**;
- personal data revealing **trade union membership**;
- **genetic data**;
- **biometric data** (where used for identification purposes);
- data concerning **health**;
- data concerning a person's **sex life**; and
- data concerning a person's **sexual orientation**.

The ICO guidance confirms that "special category data includes personal data **revealing or concerning** the above types of data. Therefore, if you have inferred or guessed details about someone which fall into one of the above categories, this data may count as special category data. It depends on how certain that inference is, and whether you are deliberately drawing that inference."

Additional conditions must be met in order to process special category personal data:

- Processing of special category personal data is seen as more intrusive than processing other kinds of personal data and may require a higher level of security.
- You can only process special category personal data if you meet one of the conditions set out in Article 9 of the UK GDPR. Article 9 (2)(j) covers the use of special category personal data for research purposes.
- A Data Protection Impact Assessment (DPIA) is more likely to be required.
- The University has a policy for processing special category personal data and criminal offence data. *Please note that this policy will be reviewed and updated in 2022/23.*

What is criminal offence data?

The UK GDPR gives extra protection to "personal data relating to criminal convictions and offences or related security measures". This is referred to as criminal offence data and covers a wide range of information about offenders or suspected offenders:

- criminal activity
- allegations
- investigations
- proceedings
- penalties
- conditions or restrictions placed on an individual as part of the criminal justice process
- civil measures which may lead to a criminal penalty if not adhered to.

It may also include personal data about:

- unproven allegations
- information relating to the absence of convictions.

It does not cover information about other individuals, including victims and witnesses of crime, but personal data about these individuals is likely to be sensitive and to require particular care.

Additional conditions must be met in order to process criminal offence data:

- Processing of criminal offence data is seen as more intrusive than processing other kinds of personal data and may require a higher level of security.
- You can only process criminal offence data if you meet one of the conditions set out in Schedule 1 of the Data Protection Act 2018. There is a condition at paragraph 4 for research purposes.
- A Data Protection Impact Assessment (DPIA) is more likely to be required.
- The University has a policy for processing special category personal data and criminal offence data. *Please note that this policy will be reviewed and updated in 2022/23.*

ICO Guidance:

Personal data - https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/

Special category personal data - https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/

Criminal Offence Data - https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/criminal-offence-data/

The ICO is currently developing some further guidance on **anonymisation and pseudonymisation**. This University guidance will be updated when the ICO guidance is finalised. In the meantime, the guidance on anonymisation under the Data Protection Act 1998 is still helpful: Anonymisation: managing data protection risk code of practice (ico.org.uk).

Further Information and Support: Information Governance Team

Sharepoint: https://sheffieldhallam.sharepoint.com/sites/3037/SitePages/Information%20Governance.aspx

Email: DPO@shu.ac.uk